DOCUMENT RESUME

ED 106 364 TM 004 482

AUTHOR Impara, James C.
TITLE Determining Assessment Content--Meeting Real Needs.
PUB DATE [Apr 75]
NOTE 14p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, D.C., March 30-April 3, 1975)

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS Credibility; *Educational Assessment; Educational Change; Evaluation; *Evaluation Criteria; Evaluation Methods; *Evaluation Needs; *Objectives; Program Costs; Relevance (Education); *State Programs
IDENTIFIERS Oregon Assessment Program

ABSTRACT

The determination of assessment content is often made on the basis of cost, political "clout", and relevance, in that order. Three areas of assessment content are discussed: Broad areas, specific areas to be measured, and non-test information. The broad areas and non-test information are policy issues, while the determination of specific outcomes is a more technical one. Several criteria are suggested for aiding policy makers in determining broad areas and the non-test information which is to be included. The issue of determining specific outcomes may occur at the initial planning stages of assessment or as suggested by Dyer, following an assessment "trial" run. In either case the specific outcomes should be determined on the basis of the involvement of professional educators and nonprofessional educators. Oregon's methodology for determining specific outcomes for assessment is included. (Author/BJG)

ED106364

DETERMINING ASSESSMENT CONTENT - MEETING REAL NEEDS

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRO
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIGIN
ATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRE
SENT OFFICIAL NATIONAL INSTITUTE OF
EDUCATION POSITION OR POLICY

Presented by:

James C. Impara, Director, Statewide Assessment
Oregon State Department of Education
Salem, Oregon

to

American Educational Research Association
Annual Meeting
March 30-April 3, 1975

TM 004 482

In each of the fifty states some sort of statewide assessment program exists. Each of these states has made a decision about what to assess. It might appear that this paper is somewhat late in terms of providing alternative strategies for determining assessment content. However, content areas to be assessed do change and new strategies become necessary. This paper attempts to address the problem of meeting real needs in determining assessment content; it is not directed toward the mechanisms for determining content.

According to Womer, "The most logical starting point for any large scale assessment is to determine its prime audiences and their most crucial needs for student outcomes information. Such a focused program probably will be useful to others outside the prime audience, whereas an unfocused program may not be of any use to anyone." (Womer, p. 16) Clearly, in order to meet needs, it must be determined whose needs will be met. The primary audience for assessment is usually either teachers, local decision-makers, or state level decision makers. On occasion an attempt is made to make more than one of these groups a combined primary audience. The difference between these alternatives usually involves the unit of analysis and the nature of the reports, although some content differences may also be indicated.

Assessment programs have begun with little thought to either the purposes of the assessment or to its audiences. In some cases assessment is begun by legislative mandate, in other cases it is the decision of the Chief State School Officer or the State Board of Education. Rarely is it initiated by local school districts or by teachers. In fact, this author is aware of no such instance where the later group has been the initiator of statewide assessment. Someone however, must make the decision about the primary focus and purpose of statewide assessment. This decision requires that those who set assessment policy know what type of information

they want or need. This can be a problem when those in charge of assessment policy do not represent the audience for whom assessment data are intended.

It is not feasible to differentiate the purposes of assessment from the audiences to be served by the assessment. If the policy makers for assessment are not representative of the prime audience then some mechanism must exist for enlightening the policy makers as to the needs of the audience. Whether or not the primary audience is represented in making assessment policy, there are two crucial aspects which impinge on the potential or actual utility of the data for meeting real needs. These two aspects are (1) trust and (2) involvement. These two conditions are "necessary but not sufficient," i.e. they do not guarentee that assessment data will be used, but without these conditions there is a greater likelihood that any assessment program will become a "shelf" program. (A "shelf" program is one which serves no useful purpose other than generating reports, print-outs, etc. that collect dust upon a shelf.)

Although most assessment programs have begun with reading or math, some have not followed this trend. Even so, cognitive outcomes are most commonly the primary focus. A few states have implemented assessment in the affective domain, while even fewer have assessed psychomotor skills. The content of assessment can be considered to include several dimensions:

(1) The broad areas upon which assessment is focused, such as reading, math, citizenship, or self concept.

(2) The specific outcomes within these areas.

(3) The non-test information used to help explain or to categorize assessment results, such as sex of student, parents educational level, school size, or expenditures per pupil.

Determining the broad areas and the non-test information are clearly policy matters, while selecting the specific outcomes is a more technical issue. Both of

the policy concerns can be (and have been) resolved in various ways. The decision on the broad areas is often made by legislation, as in Florida, or by the Chief State School Officer or by an "Advisory Committee" as in Oregon. The decision is often based upon input received through "happenings" such as town hall meetings, Delphi techniques, public opinion polls, or discussions (formal or informal) with key legislatures or key school district superintendents. Regardless of how the decision is made, a number of factors are usually considered:

(1) The estimated costs of the program, in particular the costs required for acquiring or developing the test as a function of content.

(2) The political impact of the broad area to be assessed. That is, how much "clout" will the published results have in bringing about needed instructional improvement or redistribution of resources.

(3) The relevance of the results to the primary assessment audience. (In some cases "clout" and relevance are the same depending on the area and the audience.)

It is not unusual for these three considerations to occur in the order in which they are listed. Thus, in considering assessment content the relevance of the results is often not the foremost concern. Under these conditions how then can we determine assessment content which will meet real needs.

If the primary audience is to be the classroom teacher, then statewide assessment probably can not do an effective job of meeting their needs. In fact, this author is becoming more and more convinced that this is a correct point of view. That is not to say that classroom teachers can not be users of assessment data. It is to say that teachers should not be the primary assessment audience. Their needs may best be met by local school district assessment or testing programs. The rationale for this statement is that the needs of teachers are simply too diverse to be met by a statewide assessment program. If not teachers, who then should be the primary

audience? I believe that the focus of assessment should be state-level decision-makers.

No matter which state-level audience is determined to be the primary one, the results will still have an impact (if these results are reported in a useful way) on the operation of local schools. In some instances, the impact may be very direct and immediate, in other instances, more indirect and over a longer term. What then are the implications for this position for determining assessment content? One clear implication is that whatever is measured should be almost universally accepted as being important, or should be "visible" enough to justify the high cost of assessment. Another implication is that the credibility of the results must be high, i.e. high trust level must exist between the assessment program, and the primary audience, and for those who will be affected by the decisions or actions of this primary audience. While these implications may seem trivial or at least obvious, they present some interesting dilemmas for deciding on both the general areas to be assessed, as well as the specific content. These dilemmas can only be resolved when the specific purposes of assessment are known and agreed upon by the policy makers.

Depending on the nature of the intended decisions or actions, other implications are also apparent and they relate to issues such as the type of test (mastery or differential), and data collection methodology (sampling or census), the type of report (interpretive or not interpretive). These issues are discussed elsewhere and are not considered in this presentation.

It is generally agreed that each of the dimensions of assessment content (broad areas, specific outcomes to be measure, and non-test information) should be determined at the early stages of assessment planning prior to any testing. If we assume that statewide assessment is not a one-shot affair, but will continue

on a long term basis, then we must further assume that the specific objectives of assessment will change periodically. But will the broad areas change over time, and will the non-test data change over time? Probably not to any large extent, although areas may be added and additional non-test information may be requested, (the general trend in data collection is not to delete or purge information rather it is to ask for more information). Thus, the issues of which broad areas will be of most relevance to the policy-makers and which non-test information will provide this audience with useful information for decision making must be considered early. While costs and "clout" are important factors, the relevance of the assessment results should not be overlooked in making decisions on broad areas and non-test information.

The following criteria are suggested for making the determination of the broad areas to be assessed when given student performance information:

(1) Will it be sufficient to meet the "public's right-to-know" about the status and condition of pupil performance in the public (and/or nonpublic) schools?

(2) Is there an action or set of actions which can be taken to bring about needed change?

The change may be in relation to resource allocation, or to a different emphasis on instruction in the area.

The non-test information should also be examined against these criteria. There are many items of non-test information which can be collected. Some of these items are correlated with pupil achievement. However, is it feasible to make changes based upon this information? Sometimes the answer may be yes even though the policy-makers have no control over the condition or classification. For example, policy-makers can not control a student's race, but they may control resources which can be differentially allocated to students who are members of a particular race, or to schools with a high concentration of students of a particular race.

These criteria in themselves do not suggest what assessment content ought to be. They represent instead a test for insuring that needs can be met. When deciding whether assessment should focus on cognitive outcomes, psychomotor outcomes, or affective outcomes, these criteria become quite relevant. After a preliminary decision has been made several questions must be answered by the policy-makers. These questions include:

(1) If change is needed, how long will it take for change to occur?

(2) If we know how to bring about change, can we afford the costs of doing so, or if we do not know how to bring about change, can we afford the cost of finding out?

(3) Does the public desire such change?

While these three questions are important they may on occasion be ignored if the need to meet the public's right-to-know is sufficiently strong.

By what methods should we determine which broad areas to test using these criteria? If the areas are not mandated, (which is a frequent occurance) then the assessment staff must undertake to prepare a case for each of the alternatives which can be afforded (assuming that budget limitations of the assessment program are specified). This also holds true for determining the non-test information which is to be collected. If it is not mandated, then there are several resources one can use in formulating various cases for the non-test information (e.g. Bryant et. al., 19[illegible] SUNY, 1972).

How does one determine the specific outcomes which are to be measured in assessment? It is this area which requires the maximum of involvement of those who will be affected by the assessment. The specific outcomes to be measured should be determined both by professional educators and by nonprofessional educators. (A nonprofessional educator is any one who is not employed in the field of education. This group is often called the "general public" or the "lay community.) Womer suggests that the determination of the specific outcomes should be accomplished in the early

planning stages of assessment. He goes on to say that "...subject matter objectives are not (should not be) the sole property of subject matter specialists." (Womer, p. 56) This author agrees with that preference since even though the teacher is vitally concerned with what should be learned, so is the student, the parent, board of education members, and other tax payers. The process of surveying teachers and others is a long and arduous one. To obtain meaningful input requires a great deal of time and effort on behalf of all of those who are involved in the assessment. Those individuals include not only the staff at the state level, but also those in local districts who are assisting in the determination of selecting objectives.

While Womer suggested that the determination of the objectives to be assessed should be accomplished in the planning stages, Dyer, on the other hand offers an alternative approach. He suggests that a case can be made "for the proposition that you are more likely to get useful agreement on what the outcomes of education ought to be after you have made some assessment of what the outcomes of education actually are." (Dyer, p. 23) Unfortunately, Dyer does not provide a methodology for undertaking this alternative. The argument in favor of this approach does have possibilities since in the initial assessment of an area it may be extremely difficult to establish a proper mental "set" for determining the outcomes of schooling. It is difficult sometimes to predict the consequences of an assessment and therefore it may be highly desireable to modify what is measured after some experience in interpretation and utilization of results has occured. This may not require a change in the procedures for determining assessment objectives in the initial stages of assessment. It does require however, that a high level of trust exists, so that necessary changes can be made once the assessment cycle has begun.

SUMMARY

The determination of assessment content is often made on the basis of costs, political "clout", and relevance, in that order. Three areas of assessment content are discussed: Broad areas, specific areas to be measured, and non-test information. The broad areas and non-test information are policy issues, while the determination of specific outcomes is a more technical one. Several criteria are suggested for aiding policy makers in determining broad areas and the non-test information which is to be included. The issue of determining specific outcomes may occur at the initial planning stages of assessment or as suggested by Dyer, following an assessment "trial" run. In either case the specific outcomes should be determined on the basis of the involvement of professional educators and nonprofessional educators.

OREGON'S METHODOLOGY FOR DETERMINING SPECIFIC OUTCOMES FOR ASSESSMENT

The selection of performance indicators (which in another context might be called objectives) includes several components. The initial step is to form a technical advisory committee which consists of classroom teachers and subject matter specialists who have a reputation in the State for being highly qualified in the subject area to be tested. This group is brought together to accomplish several purposes:

(1) To define in general terms the subject matter to be assessed.

(2) To set up appropriate subdivisions (domains) within the scope of the definition. For example in mathematics these broad areas might include geometry, arithmetic, measurement, problem solving, and probability and statistics.

(3) To recommend existing sources of performance indicators. These sources often include the Instructional Objectives Exchange, National Assessment, and local education agency indicators.

(4) To review the collection of performance indicators to insure that they meet a pre-established set of criteria. These criteria include statements pertaining to relevance, appropriateness for grade level, appropriateness to domain, and clarity of content (understandability).

(5) Make a preliminary selection of those performance indicators which in the review panel's opinion, are the most important and essential for student success in everyday living.

(6) To make the final decision on any performance indicators which have a "tie score" as a result of the field selection process.

The second component involves the distribution of a preliminary set of performance indicators to both professional and nonprofessional educators around the state. This step is accomplished subsequent to item 5 above. These performance indicators are distributed to "assessment coordinators" employed by each county (either county unit or intermediate education district) who are assigned by the superintendent certain responsibilities in the area of statewide assessment. Among these responsibilities is that of assisting in the selection of the performance

indicators to be measured in statewide assessment.

The coordinators are asked to have the performance indicators reviewed by local individuals who are professional or nonprofessional educators. The coordinators then obtain, by whatever means is most convenient, a consensus opinion from the participants on the indicators which should be measured in the statewide assessment program.

The number of indicators to select is given to prevent all those indicators sent out from being selected. This number is usually between one-fourth and one-half of the total sent out. The determination of the number of performance indicators is based upon an estimate of 5-6 items per indicator and a desired testing time appropriate to the grade level which is to be assessed. Thus, if fourth grade was to be assessed using a 90 minute (two session) test, and if the average time per item was estimated at one minute, then 15 performance indicators would be selected.

The criteria the state will use for determining if an indicator is selected is given at the same time as the indicators are distributed to the local assessment coordinator. This criteria is usually a combination of various weighting schemes. For example, each unit is awarded a weight of one, for one portion of the scale and for another portion is weighted according to the proportion of the state's children which are included in that unit. If a unit has 10% of the State's children, it would be awarded a weight of .5. In this example each performance indicator selected must have had an absolute score equal to one-half of the number of units responding and it must have had a population weighted score in excess of 50. Thus units which have only a small percentage of the State's population can not band together and force an indicator to be selected while the units which have a large portion of the State's population can not band to insure the inclusion of a

particular indicator. Each indicator must be selected by both a large number of units (at least half) and these units must represent a large portion of the state's population (at least half) .

The final component involves tabulating the consensus from each unit and if necessary having the advisory panel resolve tie scores. The tie breaking process is only undertaken if the maximum number of performance indicators desired would be exceeded by including all of the indicators for which ties have occurred.

Thus far it has not been necessary to develop from scratch any performance indicators for the Oregon Assessment program. This is because reading and math are the first areas to be assessed, and upon examination of existing performance indicators, there has been little need to develop new performance indicators. This condition is not expected to continue since the next area for assessment is citizenship. The advisory panel for this area is now being formed, and there is an expectation that development of some performance indicators will be necessary.

REFERENCES

Bryant, Edward C., Glaser, Ezra; Hanson, Morris; and Kirsh, Arthur. Associations Between Educational Outcomes and Background Variables: A Review of Selected Literature, NAEP. 1974.

Dyer, Henry. "The State of the State Assessment Art," Speaking Out: Law, Education and Politics. Proceedings of the Invitational Conference on Educational Assessment and Educational Policy. ETS. 1974.

Variables Related to Student Performance and Resource Allocation Decisions at the School District Level. The University of the State of New York, State Education Department, Bureau of School Programs Evaluation. 1973.

Womer, Frank. Developing a Large Scale Assessment Program. Cooperative Accountability Project, Denver, Colorado. 1973.